Since the distributional assumptions underlying classical parametric statistical tests are often violated (by, for example, extreme skew or kurtosis in the data), it is often necessary to use non-parametric methods instead. For example, the rank-sum test (Mann-Whitney "U") is an alternative to the classical t-test that, unlike the t-test, makes no assumptions about how the underlying data are distributed. The rank-sum test examines whether, over all, the members of one group tend to rank higher than the members of another group. Because it is based on ranks rather than means and standard errors, the rank-sum test is much more robust than the t-test where distributions are skewed and/or outliers are present.

But of course, it isn't very useful to know that there is a significant difference between the two groups, if you don't know how big the difference between the groups actually is. Indeed, a common complaint about nonparametric statistics is that they allow you to assess statistical significance but they don't let you measure effect size. Fortunately, there are techniques for measuring the size of an effect that are insensitive to distributional nastiness (like skew, kurtosis, or outliers) just like nonparametric statistical tests are. One class of such methods are called Hodges-Lehmann estimators.

The Hodges-Lehmann estimator $\hat{\Delta}$ for the difference between two groups provides a good illustration. If one group is the x_i's: $x_1, x_2, x_3, \ldots x_n$ and the other group is the y_j's: $y_1, y_2, y_3, \ldots y_m$, then the Hodges-Lehmann estimator for the difference between the x's and y's is determined as follows.

1. Calculate the difference between every possible pair of x's and y's: $d_{i,j} = y_j - x_i$. There will be a total of *m* times *n* such differences.
2. Rank the list of these differences in ascending order.
3. Pick the median from this list: $\hat{\Delta} = \text{median}\,\{y_j - x_i\}$

That's all there is to it. Now, what does the Hodges-Lehmann estimator do? It does two things. First, it is the best unbiased estimator of the median of the distribution of possible differences between the median of x and the median of y. Second (and less technically), if you increase each x_i or decrease each y_j by an amount $\hat{\Delta}$, you will eliminate the difference between the x's and y's as seen by the rank-sum test.

Since the Hodges-Lehmann estimator $\hat{\Delta}$ is the median of a distribution (the distribution, that is, of $d_{i,j}$'s), you can estimate the confidence limits for $\hat{\Delta}$ similarly to the way you estimate the confidence limits of the median of a set of measurements (but with a degrees-of-freedom correction, since there are far more $d_{i,j}$'s than there are original x_i's and y_j's). Here's how you do it:

1. Decide on a confidence level, and the probability $\alpha/2$ that should lie in each tail.
2. Look up $Z_{\alpha/2}$ for the confidence level you've chosen.
3. Compute the rank of the lower confidence limit as $R_L = \frac{mn}{2} - Z_{\alpha/2}\sqrt{\frac{mn(m+n+1)}{12}}$
4. If R_L is a non-integer, round down.
5. Compute the rank of the upper confidence limit as $R_U = mn - R_L + 1$
6. Look up the values that correspond to the ranks R_L and R_U in your ranked list of the $d_{i,j}$. These are the confidence limits for $\hat{\Delta}$.

If *m* and *n* are small (say, *mn*<100 or so), then other methods are necessary to estimate the confidence limits; see Helsel and Hirsch (or a good nonparametric statistics text) for details.

For further reference:

Helsel, D. R. and R. M. Hirsch, *Statistical Methods in Water Resources*, 522 pp., Elsevier, New York, 1992.